CRM D0008914.A2 / Final
October 2003

Statistical Analysis of USMC Accidental Deaths

W. Brent Boning • Michael D. Bowes •
Catherine M. Hiatt

4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Approved for distribution: October 2003

Alan J. Marcus, Director
Infrastructure and Readiness Team
Resource Analysis Division

Contents

This page left intentionally blank.

Summary

Background

The United States Marine Corps has an interest in preventing accidental deaths. One aspect of developing an effective program is to understand how individual characteristics and events may be associated with deaths. The objective of this study is to determine in an analytically sound manner, the variety of factors that explain fatality rates. The focus is on non-combat ground fatalities, including motor vehicle accidents that account for a majority of deaths.[1] The information developed can be used to tailor safety messages to the points in a Marine's career where risk is high. Further, it may allow those in direct leadership positions to intervene proactively to reduce fatalities among those identified as most at risk.

Tasking and study approach

We addressed the following tasks:

- *Build a data set.* We built a data set characterizing individual Marines and fatalities covering the time period from June 1996 through March 2003. Although personnel records are the principal source of data, we also incorporated additional information from safety records.
- *Perform statistical analysis.* We undertook a statistical analysis to estimate the risk of fatality. In this analysis we characterized the relationship of various individual characteristics and career events to risk. We used an approach that is frequently used in epidemiological studies and is referred to as survival analysis or

1. Ground fatalities include those resulting from on-duty mishaps (except operational aviation), vehicle accidents, and off-duty recreation.

hazard rate modeling [1, 2]. In addition, we provide a summary analysis of Marine Corps fatality data that includes some comparisons against equivalent civilian data.

- *Predicting future fatalities.* Using estimates developed to explain individual risk, we provide predictions of future fatalities by quarter.

Summary of study findings

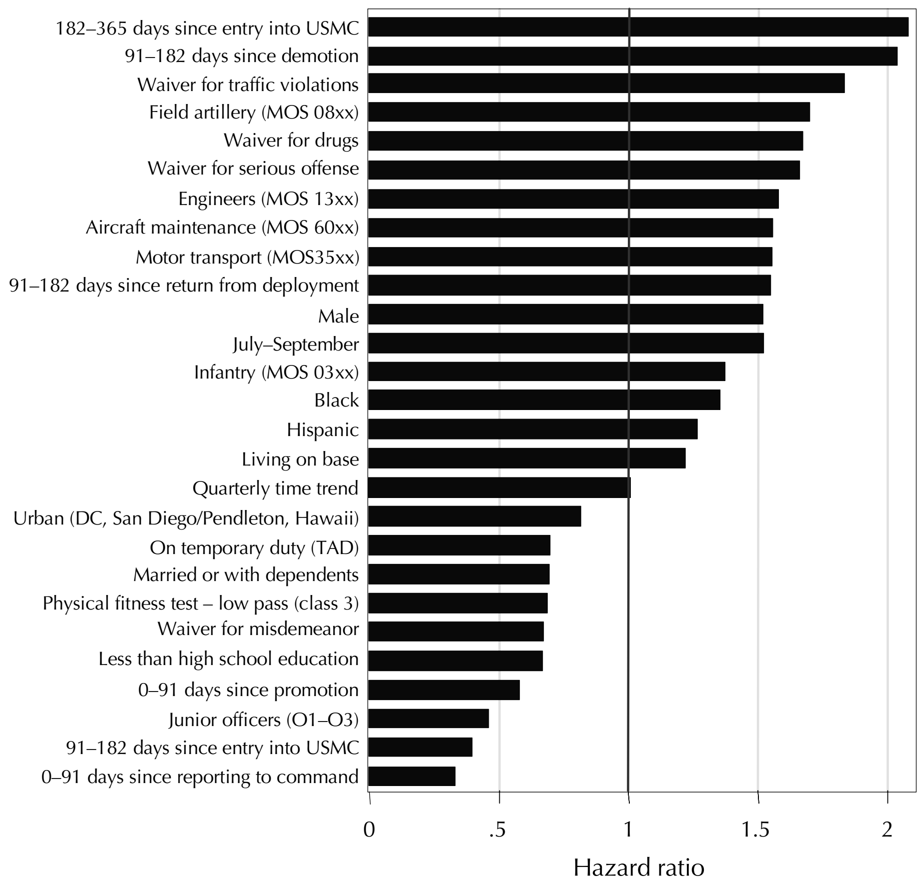
Our primary interest is in the factors that explain fatality rates. Some of our results are summarized as follows:

- *Early career.* During the last half of the first year in service, a Marine's risk of death is twice what it is at other times.
- *Post deployment.* The period after deployment is associated with substantial risk of fatality. The risk of vehicle-related fatalities is particularly high (about twice the normal risk) during this period.
- *Demotion.* In the period 3 to 6 month after a demotion, the risk of accidental death is twice as high as at other times.
- *MOS groups.* Artillery, aviation mechanics, motor transport mechanics, engineers, and infantry MOSs have risks of accidental death that are 37- to 69-percent higher than that of other Marines.
- *Location.* Marines living on base have a 21-percent higher risk than others. Marines who are not based in urban areas face a 23-percent greater risk than those in urban areas.
- *Enlistment waivers.* Individuals with a history of drug use, traffic violations, or serious offenses prior to enlistment have a 66- to 83-percent increase in risk as compared to those who enter without waivers.
- *Race and ethnicity.* There is a strong relationship between race or ethnicity and risk. Other things being equal, blacks have a 35-percent and Hispanics a 26-percent greater risk of fatality than do whites.

- *Time of year.* Late summer months are associated with a 52-percent increase in risk relative to the winter quarter.

The statistically significant results are illustrated in figure 1. The figure is based on all accidental ground fatalities. (In the paper, we take a separate look at vehicle-related fatalities). A hazard ratio of greater than one indicates higher risk. For example, a value of 1.5 means a 50-percent higher risk. A value of less than one means lower risk.

Figure 1. Risk of fatal accident associated with various characteristics



Some other findings are as follows:

- *Vehicle fatalities by time of day.* In looking at vehicle fatalities, the most striking feature of the data is an unusually high death rate during early morning rush hours. This may suggest that Marines are taking extraordinary risks to make muster.

- *Comparison to civilian fatality rates.* In general, Marine accidental death rates are below those for civilian males of equivalent age.

Organization of this report

In the first section, we present an overview of USMC accidental deaths statistics and provide some comparison to civilian rates. In the second section, we describe our statistical model and results.

Summary statistics on USMC accidental deaths

In this section, we present information on our data sources. In addition, we provide summary information on USMC accidental deaths from calendar years 1997 through 2002. The analysis we present here will provide a general sense of trends and introduce some of the concerns that we will explore more carefully in later analyses. Our focus is on ground fatalities that result from accidents unrelated to combat. We do not consider operational aviation incidents.

Data

The data we use combine information from four sources. The Naval Safety Center maintains the formal records on USMC accidental deaths, but because those records offer little of the demographic information we need to identify individual risk factors, we rely for the most part, on information from other Marine Corps sources. For basic demographic and career data, we draw on the Headquarter Master Files (HMF). These files provide a snapshot of USMC personnel at the end of each quarter. The data include both demographic characteristics (e.g., birth date, sex, race, and marital status) and career information (rank, MOS, PMCC, and deployment). We combine this with data from the ARSTAT files to capture information on losses (retirement, end of enlistment, and death). We also draw on CNA's Street-to-Fleet file, which records accession data. This file provides us with information on waivers (traffic, alcohol, or drug offenses) granted during enlistment that might be related to risk of fatality.

Although the personnel files do include information on fatalities, the records are not entirely adequate. For example, no distinction is made between accidental deaths and suicide. Another problem comes from the practice of discharging Marines facing imminent death under a temporary disability retirement (this allows additional benefits for family members). Such cases are not distinguished from other disability retirements. To address these data problems, we matched Safety Center records to specific individuals in the personnel records. Using the limited personal information that is available in the Safety Center files (age, sex, rank, and MOS), we were able to identify most fatalities. In doing so, we were able to exclude suicides and combat-related deaths and include most of the accidental deaths that were recorded as disability retirements.[2] The merging of the Safety Center data gave us some additional details on the circumstances of death. For our later analysis, we use data covering the identified fatalities and a representative sample (5 percent) of all Marines serving between June 1996 and March 2003.

Summary statistics on deaths

Numbers, trends, and predictions

Our data include 455 accidental deaths in the USMC between CY 1997 and CY 2002. Figure 2 shows the number of accidental deaths and the fatality rate per 100,000 Marines by year. The fatality rate for Marines compares favorably to that of a comparable civilian population. In 2000, the accidental death rate was 55 per 100,000 for U.S. males between the ages of 17 and 29.[3] For male Marines in this same age group, the average fatality rate over the 5 year period was 49. Because the Marine population is younger and more heavily male than the U.S. population as a whole, two factors that are associated with higher accidental death rates, any comparison that does not adjust for the mix is biased against the USMC.

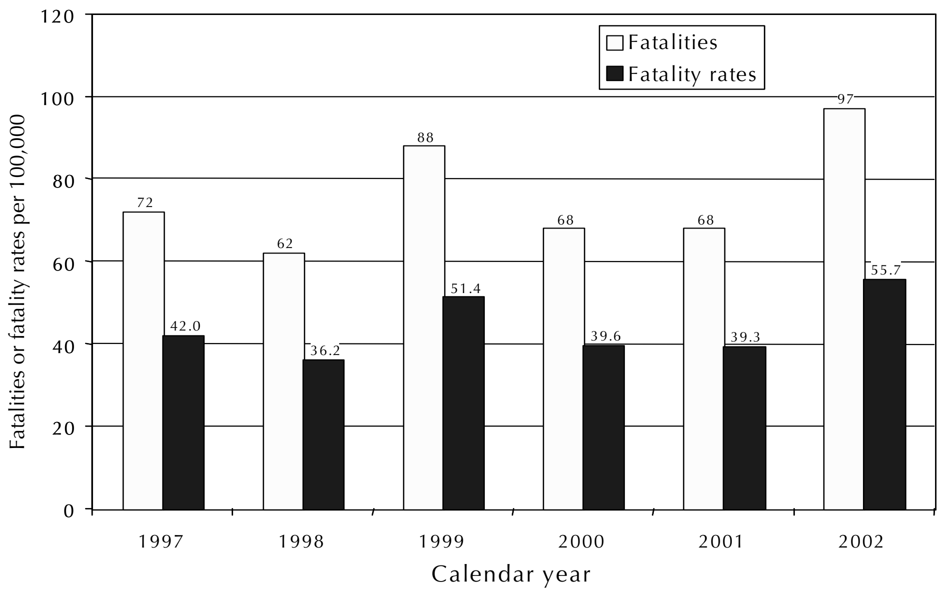
Table 1 shows the number of deaths annually by type of accidental death. Vehicle-related deaths listed here include deaths involving private vehicles, including pedestrian deaths. An additional 25 vehicle-related deaths are included in the operational category. Overall, more than 70 percent of the USMC fatalities are vehicle-related.

---

2. There were nine deaths in the Safety Center data that we could not confidently match to personnel records or otherwise explain.

3. Civilian rates are derived from the National Center for Injury Prevention and Control, *Leading Causes of Death Reports.*

Figure 2. Trends in Marine Corps accidental ground fatalities[a]



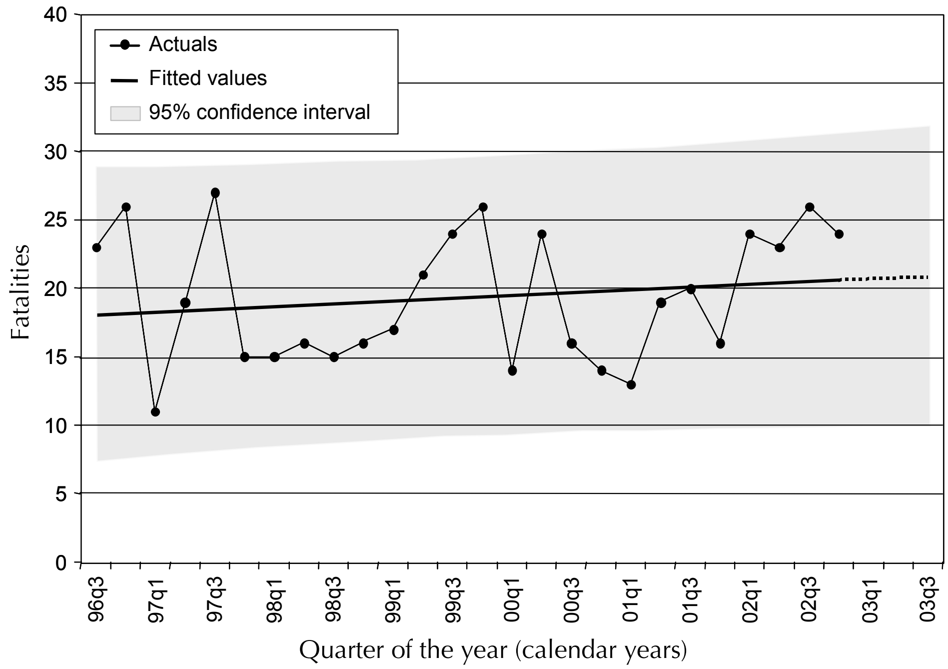
a. The number of Marines used to calculate fatality rates is derived from September HMF files for each year.

Table 1. Accidental deaths annually, by cause, 1997–2002

Calendar year	Private vehicle	Operational	Off-duty recreation	Total
1997	60	8	4	72
1998	42	13	7	62
1999	63	18	7	88
2000	48	14	6	68
2001	39	19	10	68
2002	65	21	11	97
Total	317	93	45	455

Figure 3 shows quarterly fatality data and a simple linear regression fit of these data. Behind the obvious variability in the number of fatalities over quarters, there appears to be a modest upward trend in fatalities over time (0.4 additional per year). The data in table 1 suggest that increases in operational and off-duty deaths may be behind this trend. The shaded area in figure 3 represents a 95-percent confidence interval on observed fatalities. An observation outside this range might be considered a significant departure from expectation.

Figure 3. Quarterly fatality data with a linear regression fit

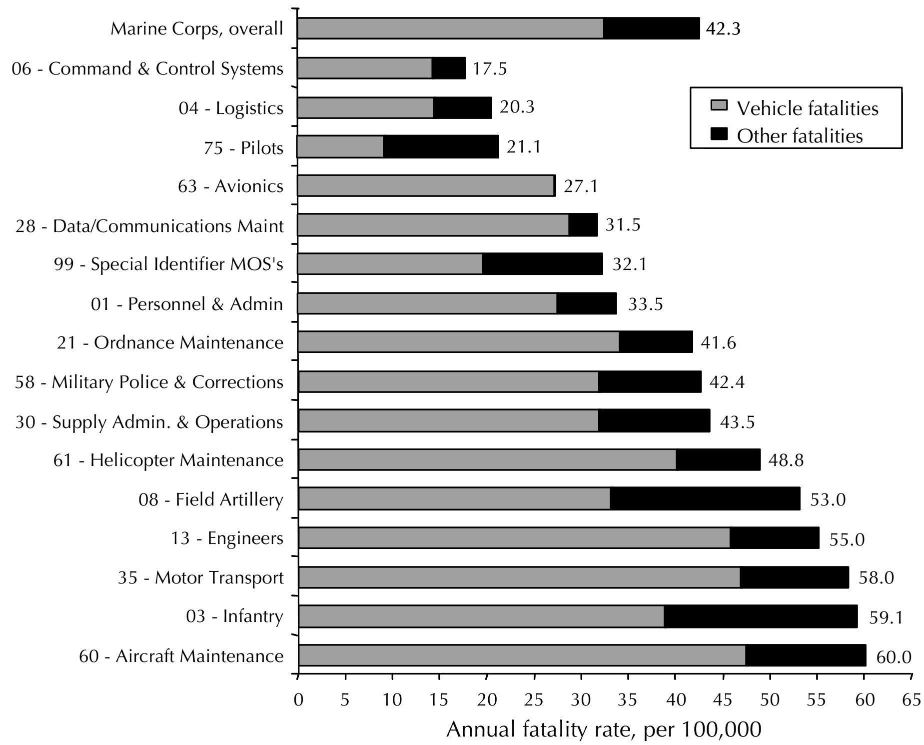


Similar regression approaches using aggregate fatality data or fatality rates can be used to predict future deaths. The Naval Safety Center does so, using an approach that accounts for holiday weekends and time of year. Such predictions are useful for the safety message they convey. However, they are limited in predicting fatalities in situations where significant changes in underlying risk factors occur. We will address the prediction issue later, demonstrating an approach based on individual data and risk.

Fatality rates by MOS

There are pronounced differences in the ground fatality rates among military occupational specialties (MOSs). Figure 4 shows the fatality rates for the 16 largest MOS groups (at the two-digit level). These groups comprise 75 percent of the Marine Corps population. Both the highest overall accidental death rate and the highest vehicle-related death rate are in aircraft maintenance (MOS 60xx). This total is almost matched by the infantry (MOS 03xx), though a lower percentage of infantry deaths are vehicle related. Other MOSs with high death rates are motor transport (35xx), engineers (13xx), field artillery (08xx), and helicopter maintenance (61xx). The differences in death rates across occupations might indicate that some MOSs draw people who are more likely to take risks or that certain jobs present the opportunity for greater risk. We explore this possibility in more detail later.

Figure 4. Average annual fatality rates by MOS[a]



a. Fatality rates based on USMC accidental deaths from 1997–2002. For convenience, second quarter 2003 workforce by MOS is used as the divisor. Vehicle fatalites include both personal and operational vehicle fatalities.

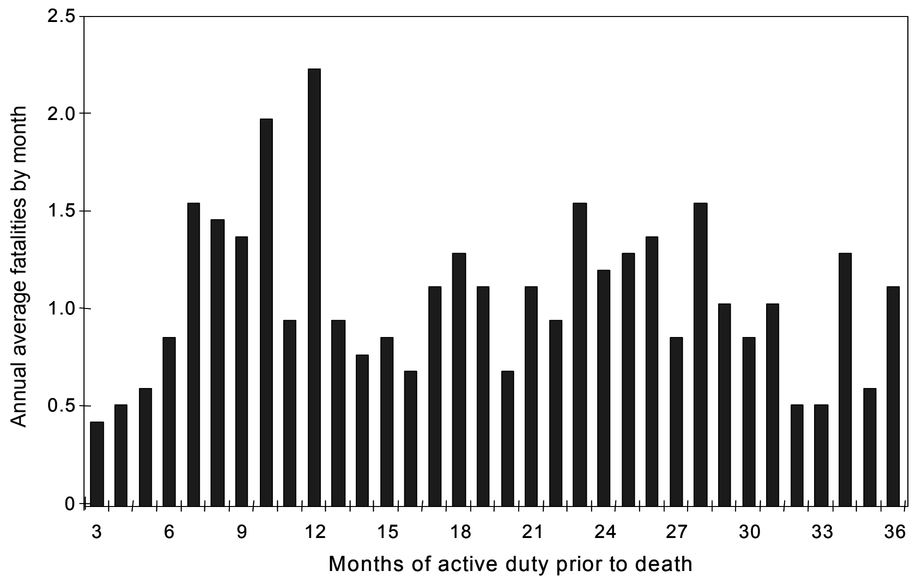
Fatalities by time in service

Frequency of death varies by duration of service. Figure 5 presents the average number of historical fatalities for an annual cohort of new Marines, by months of service until death. The data here are for Marines with active-duty base dates between July 1988 and June 2000[4]

4. We discard more recent data in order to avoid misrepresenting the distribution (for Marines with recent entry dates, it is not possible to have observed all deaths that may eventually occur). Data from earlier years is used to help smooth the distribution.

and are based on vehicle-related deaths identified in HMF files. We omit the first 2 months because the personnel records do not include some individuals who die during the first 2 months of duty.

Figure 5. Frequency of accidental death by duration of service[a]



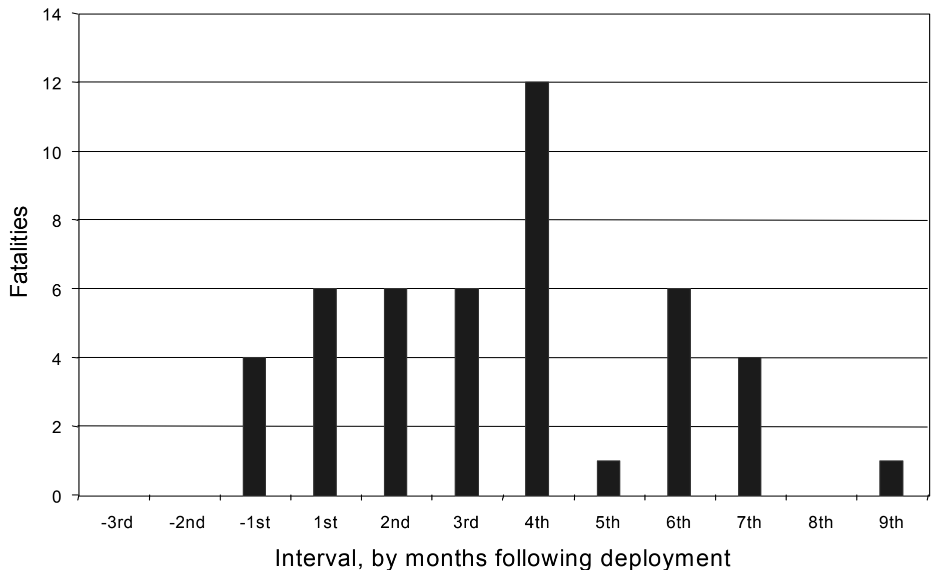
a. Average annual fatalities based on Marines entering service between July 1988 and June 2000.

The figure gives some sense of the profile of risk by duration of service. It appears that risk is relatively low early in a career, perhaps due to stringent restrictions on liberty. At about 7 months, there is a sizable jump, possibly associated with a loosening of restrictions. Risk remains high throughout the remainder of the first year (except for an unusual drop at 11 months). Following this, the rate is lower, until about the end of the second year, when there is a modest jump. That increase is perhaps associated with return from deployment. The techniques we present in the next section are explicitly designed to explain survival time data like that depicted in figure 5. These techniques will allow us determine the individual factors that best explain the frequency and timing of death.

Fatalities following deployment

In figure 6, we show frequency of death following deployment. There appears to be a period of high risk in the first 6 months after a return from deployment. We see, for example, 6 deaths during the first month back from deployment; the number of fatalities then peaks at 12 in the fourth month back.

Figure 6. Frequency of accidental death relative to deployment return



a. Based on fatalities among Marines who served in the period 1997 through 2002 and who returned from deployment.

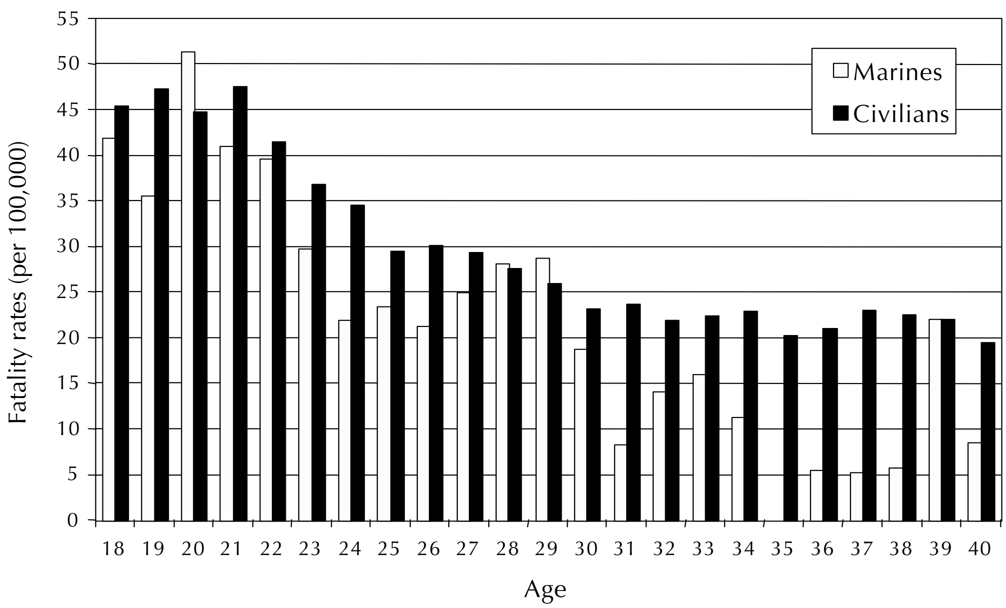
We do have some concern with the accuracy of deployment return dates. Prior to deployment, return dates are entered in the HMF files as estimates. Once deployment has actually ended, a corrected date will appear in subsequent quarterly files. However, quarterly personnel files usually do not include individuals whose service ends during the quarter (including those who die). As a result, we can't be sure we will see the correct end-of-deployment date for those who die or leave the service shortly after deployment. (In fact, we can't even be sure they deployed.) The relatively high number of deaths in the month prior to the end of deployment may reflect minor inaccuracies in listed return dates.

Motor-vehicle fatality rates by age

Given the significance of vehicle-related accidents, we took a closer at those incidents. Vehicle-related deaths typically have a strong correlation with age. This is true in the Marine Corps, as shown in figure 7.

Figure 7 shows the number of motor-vehicle fatalities per 100,000 people, by age. We show equivalent numbers for both the USMC and U.S. male population. For both civilians and Marines, motor-vehicle fatality rates are high for the 18 to 22 year olds, and then decrease with age. The Marine death rates exhibit more volatility than do the civilian death rates, with an increasing rates for 26 to 29 year olds, and then another drop-off. However, this apparent volatility might reflect a small population and the influence of one or two deaths. This caution applies even more strongly to death rates for Marines above the age of 35.

Figure 7. Annual motor-vehicle death rate by age, civilians males and Marines[a]



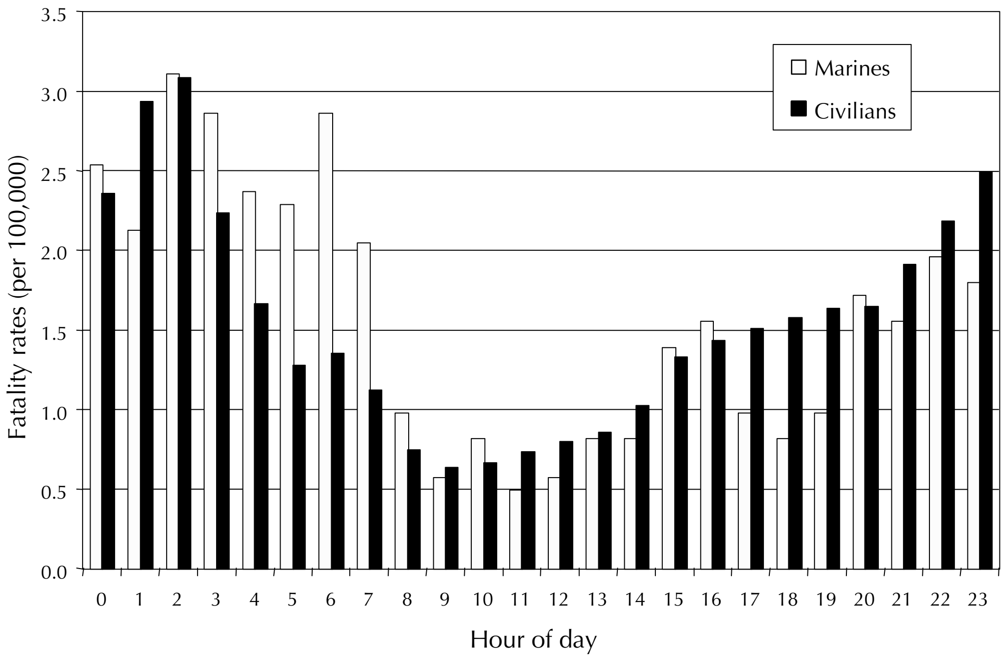
a. Civilian male rates are based on 2000 National Highway Traffic Safety AdminIstration (NHTSA) data. USMC rates are based on Naval Safety Center data from CY1997–CY2002 (mishap codes 20, 50, 60, and 70). Second-quarter 2003 population is used as the divisor for Marine rates.

Beyond the trends, it is interesting to see the relative difference in rates between civilians and Marines. In general, the Marine death rates are below those for civilian males of an equivalent age.

Motor-vehicle fatality rates by hour

The timing of motor vehicle accidents for the Marine Corps differs from those for civilians. In figure 8, we present fatality rates by hour of the day for civilian males (age 17 to 29) and Marines.

Figure 8. Vehicle-related fatality rates by hour of day, males aged 17–29[a]



a. Civilian rates are based on 2000 data. USMC rates are based on FY 1994–FY 2002 Safety Center data. The longer time period is used here to help smooth the distribution.

The most striking feature of these data is the unusually high death rate for Marines in the morning rush hours of 0500 to 0759. The rates are significantly above those of the civilians during the same hours. These rates seem to suggest that Marines may be taking extraordinary risks to ensure that they make muster.

We also see that night-time death rates for Marines peak between 0200 and 0300, and persist at high levels through the morning rush hours, whereas death rates for civilians peak between 2300 and 0200. This may mean that Marines are staying up later and that this is contributing to higher death rates, both as they head home and as they head to work.

This page left intentionally blank.

Hazard models of accidental deaths

The data discussed in the previous section provide a sense of the accidental deaths in the USMC and illustrate some trends and important comparisons. In themselves, they provide insight on what may be associated with accidents and some guidance for accident prevention programs. However, they do not allow us to unravel the many factors associated with fatal accidents. For example, though the infantry MOS has a high fatality rate, we cannot tell whether that reflects a concentration of younger single males (assuming that they are at higher risk of fatal accidents), or if there are other characteristics of infantrymen that, all else being equal, are associated with higher risk. To separate these effects, we employ hazard rate (also called survival) models.

Hazard rate models

We use hazard rate models to conduct our statistical analyses of factors associated with deaths. In this section, we describe the approach and explain why it is an appropriate technique for our analysis.

Background

In figure 5, we showed time-to-death for a typical cohort of Marines. This is the type of information we would like to be able to explain. Using typical regression methods to explain duration (time-to-death) data of this type presents a number of practical problems [1]. One difficulty is that the events and characteristics that might explain individual risk may be changing over time.

There are, however, techniques designed explicitly to deal with duration data. These techniques are used in the industrial engineering fields, where there is interest in explaining the time-to-failure of equipment. They are used in the medical fields where the interest may be in explaining survival time following treatment or diagnosis. Economists use these techniques to explain duration of unemployment. The technique is often referred to as survival analysis. In the current context, the approach is to model the probability that a particular individual will die, given that others at potential risk have survived.

Modeling assumptions and techniques[5]

The model asserts that the risk of a death occurring at time t (t in our model is time since the sixteenth birthday) for an individual j is a function of time and personal characteristics:

$$h(t) = h_0(t)\exp(x\beta)$$

This is called the hazard function. The function $h_0(t)$ describes how the baseline risk varies over time; the expression $\exp(x\beta)$ expresses how that risk increases or decreases with changes in a set of x variables that describe the characteristics of the Marine at that point in time. This particular specification means that the proportional effect of an increase in x does not depend on age. It is called a *proportional hazard model*.

The purpose of the model is to determine how the characteristics x are associated with risk. This is done using maximum likelihood estimation of the hazard functions. We estimate coefficients (β) for the variables in the model to best fit the observed data. Specifically, we select coefficients on the characteristics of Marines to maximize the probability of observing the deaths that actually occurred at each particular time (as measured from the 16th birthday). To do this, we maximize the likelihood function

$$L(\beta) = \prod_{k=1}^{K}\left(\frac{\exp(x_{jk}\beta)}{\sum_i \exp(x_{ik}\beta)}\right)$$

where, for each time (age) a fatal accident occurred (k in the equation), the numerator includes the particular Marine (j) who has died, and the denominator reflects the set of all Marines at risk at that age.

5. References [1] and [2] provide introductions to survival analysis.

In dealing with duration data, hazard rate models are preferred to alternative statistical techniques because they address the various problems that arise in the standard regression techniques. In particular:

- Hazard rate models can explicitly represent the complex stochastic process underlying survival times. The assumptions behind standard ordinary least squares, probit, and censored regression models are usually not as well suited to explaining time-to-death.
- The hazard models specifically address data-censoring (or truncation) problems. Data available will usually cover a narrow window of time. The hazard rate models account for observations that were at risk before we observed them (Marines in the data who began active duty prior to the period covered by the data) or are still at risk when we stop observing them (Marines who remain in the Marine Corps after our data ends). By addressing these concerns, hazard rate models avoid biased estimates.
- The approach can deal with time-varying characteristics. Time-to-death is likely to depend on personal characteristics and events that change over time. Designing a regression approach that would explain survival time would present a real challenge. In the hazard model, the individual's characteristics are re-evaluated at each point in time that a death occurs.
- The hazard rate models use data effectively in determining relative risks. Some means of distinguishing between propensity for death and simple population demographics is required. For example, deaths of male Marines exceed female deaths in number. This must be due in part to the fact that there are more males than females in the USMC. It may also be that males are more likely than females to be in a fatal accident. To separate these two effects, the method uses data on fatalities and comparable survivors.
- More generally, the method allows us to look systematically at complex combinations of risk factors. Although the graphical techniques of the previous section may provide insight, they become overwhelming if we try to use them to unravel the

many factors and combinations of factors that might be associated with fatal accidents.

Interpreting results

The model of interest estimates the risk of an accidental death associated with a set of demographic and career variables. Results can be expressed either as *hazard rates* or as *coefficients*. The hazard rate compares the risk for two people who are the same except for a unit difference in one particular characteristic. A hazard rate of 1 (or close to 1) indicates that the risk is not appreciably different for Marines with that characteristic than for those without. A value of less than 1 indicates lower risk. For example, a value of .5 means that an individual has only half the risk of someone without the characteristic. Similarly, values above 1 indicate higher risk. The estimation actually determines the coefficient β_i and each hazard rate is calculated as $\exp(\beta_i)$. We generally discuss hazard rates, but also provide the coefficient to help estimate risks for combinations of characteristics, as we explain later.

When interpreting the results, it is also important to note the *p-value* of the variable. The p-value indicates how sure we can be that the hazard rate differs from 1. Typically, researchers consider coefficients with p-values of less than .1 to indicate a variable that is significantly associated with a different risk. Of course, the difference in significance between a coefficient with a p-value of .099 and .101 is very small.

Data limitations and issues

Our analysis covers all Marines who served between June 1996 and March 2003, including 515 identified fatalities. The potential size of this data set presented a challenge. More than 450 thousand Marines served during the period, and there are quarterly observations on each (5.1 million observations). To make the analysis manageable, we drew a 5-percent sample to represent the Marines who were not fatalities. (All of the identified fatalities from this period are included in our analysis). This sample is weighted in the analysis to represent the full population. Actually, we used two samples. The first was used to identify variables of interest; the second was used to make final coefficient estimates. This approach avoids "data mining." We were reassured to find little difference in the estimated coefficients between the two samples. This tends to confirm that the samples are big enough to be representative of the overall population (as a look at their makeup had already suggested).

Although the data we have provide a rich basis for analyses, there are limitations. One is that data are typically recorded only while the individual is still in the Marine Corps. Our records go up to the end of the quarter prior to the individual's leaving, but they do not record an individual's status when he left. Thus, we have to assume that characteristics do not change between the last quarter observed and the date of separation from the Marine Corps.

In addition, the nature of the quarterly snapshot means that some information on changes within the quarter is lost. If a Marine finishes recruit training early in the quarter, enters a school command, and then enters another command late in the quarter, the data record only that latest command and the date that tour began. Thus, we miss some of that individual's history. This is an issue mostly with regard to boot camp. We would have liked to know exactly when an individual finished recruit training. However, we cannot assume that the start-of-tour date recorded in the quarter after boot camp is the same as the date of the end of boot camp. (We do take advantage of those dates that are available. These include latest date of rank, date current tour began, deployment end date, active duty base date, and date of attrition).

Data are also missing for many individuals who separate (either by death or other separation) before the end of their first calendar quarter in the Marine Corps. This means that we were unable to obtain complete information on Marines who die during these early months. Even if we could obtain such information, we would need data on the others who separate early in order to maintain a control group. We were unable to collect such data within the scope of this study, and instead focused on risks given that the individual makes it through the first calendar quarter.

A final caution is related to the small number of fatalities. Although we have a large sample from the Marine Corps as a whole, there are relatively few deaths (515 overall and fewer than 400 vehicle-related deaths). This means that each observation is relatively influential. Thus, the estimated risks associated with particular characteristics might have changed if a few different individuals were involved in the accidents.

Estimating risks for accidental deaths

The first model we discuss evaluates the risk of any ground-based accidental deaths. Later we look more specifically at vehicle-related deaths. The regression results are listed in table 2. We present the results in two ways. The first column is the *hazard rate*, which represents the relative risk associated with the variable. The second column presents the estimated coefficients (which are simply the logarithms of the hazard rates). We provide the coefficients to enable readers to calculate total risk. We then describe the noteworthy results. Several of the results match intuitive expectations as to who is more likely to engage in risky behavior. Tests of specification and goodness of fit are discussed in appendix A.

Table 2. Estimation results for the risk of accidental death[a]

	(1) Hazard ratio[b]	(2) Coefficient	(3) p-value
Male	1.52	0.42	0.100*
Urban (DC, San Diego/Pendleton, Hawaii)	0.81	-0.21	0.038**
Personnel and admin (MOS 01xx)	0.88	-0.13	0.589
Infantry (MOS 03xx)	1.37	0.31	0.028**
Logistics (MOS 04xx)	0.87	-0.14	0.701
Command and control systems (MOS 06xx)	1.05	0.05	0.877
Field artillery (MOS 08xx)	1.69	0.53	0.054*
Engineers (MOS 13xx)	1.57	0.46	0.030**
Ordnance (MOS 21xx)	1.34	0.30	0.341
Data/communications maintenance (MOS 28xx)	0.93	-0.07	0.819
Supply admin and operations (MOS 30xx)	1.15	0.14	0.528
Motor transport (MOS 35xx)	1.55	0.44	0.015**
Military police and corrections (MOS 58xx)	1.13	0.13	0.682
Aircraft maintenance (MOS 60xx)	1.55	0.44	0.049**
Helicopter maintenance (MOS 61xx)	1.30	0.26	0.314
Avionics (MOS 63xx)	1.02	0.02	0.951

Table 2. Estimation results for the risk of accidental death[a] (cont'd)

	(1) Hazard ratio[b]	(2) Coefficient	(3) p-value
Pilots (MOS 75xx)	2.06	0.72	0.147
Special identifier MOSs (MOS 99xx)	1.27	0.24	0.365
Hispanic	1.26	0.23	0.092*
Black	1.35	0.30	0.023**
Other non-white races	0.88	-0.13	0.574
Physical fitness test – fail (class 4)	0.82	-0.20	0.631
Physical fitness test – low pass (class 3)	0.69	-0.37	0.104
Physical fitness test – med pass (class 2)	0.95	-0.05	0.643
Junior enlisted (E1-E3)	1.17	0.15	0.239
Senior officers and enlisted (O4-O9 & E7-E9)	0.46	-0.77	0.088*
Junior officers (O1-O3)	0.60	-0.52	0.179
91–182 days since entry into USMC	0.40	-0.92	0.007***
182–365 days since entry into USMC	2.08	0.73	0.000***
0–91 days since reporting to command	0.33	-1.11	0.000***
91–182 days since reporting to command	0.80	-0.22	0.194
0–91 days since promotion	0.59	-0.54	0.001***
0–91 days since demotion	0.94	-0.06	0.875
91–182 days since demotion	2.03	0.71	0.023**
Married or with dependents	0.68	-0.38	0.001***
On temporary duty (TAD)	0.70	-0.36	0.039**
0–91 days since returning from deployment	1.33	0.29	0.268
91–182 days since returning from deployment	1.54	0.43	0.095*
Education beyond high school	0.87	-0.14	0.590
Less than a high school education	0.67	-0.41	0.068*
AFQT score	1.00	0.00	0.313
AFQT score not available	1.17	0.16	0.662
Waiver for serious offense	1.66	0.51	0.003***
Waiver for misdemeanor	0.67	-0.40	0.093*
Waiver for drugs	1.67	0.51	0.043**
Serious offense waiver not available	0.94	-0.06	0.703
Waiver for traffic violations	1.83	0.60	0.071*
Living on base	1.21	0.19	0.093*
April–June	1.32	0.28	0.062*
July–September	1.52	0.42	0.003***
October–December	1.34	0.29	0.027**
Quarterly time trend	1.01	0.01	0.078*
Wald Chi-squared test statistic	202.72		0.000***

a. * significant at 10%; ** significant at 5%; ***significant at 1%.

b. For categorical variables, hazard ratios are interpreted relative to an excluded category. For example, the Black and Hispanic values are relative to whites and MOS values are relative to individuals in MOS groups not listed.

Individual factors and risk

Males have a higher risk

Controlling for other factors, we find that male Marines are at higher risk than female Marines. This result is not surprising—accidental deaths, and particularly motor vehicle deaths, are much more common among males than females in the United States. Thus, we would expect that this characteristic would also exist in the Marine Corps. However, the magnitude of the risk may be of interest. In the Marine Corps, 97 percent of the deaths are males, and 94 percent of the population is male. The death rate for males is more than 90-percent higher than that of females, but our estimated hazard rate for males is only 52-percent higher than for females. Other differences in characteristics must account for the difference in death rates.

This coefficient provides a good example of how the estimation sorts out contributions to risk of multiple factors, while controlling for population characteristics.

Being married or having dependents lowers risk

We included a variable that represents the risk associated with being married or having other dependents. The results indicate that those who have dependents have a 32-percent lower risk than those who are not married and have no dependents. This may reflect a change in the behavior patterns associated with marriage and parenthood.

Blacks and Hispanics have higher risks

All else being equal, blacks have a 35-percent and Hispanics a 26-percent greater risk of having a fatal accident than do whites. Other groups are not significantly different from whites.

Infantry, artillery, mechanics, and engineers have higher risks

We included variables to identify the largest MOSs in the Marine Corps. Specifically, we took the 16 largest 2-digit MOSs, comprising 75 percent of the Marine Corps.[6] Among these MOS groups, motor transport and aviation mechanics (MOSs 35xx and 60xx), combat engineers (MOS 13xx), infantry (MOS 03xx), and field artillery

6. The MOS breakdown is based on the March 2003 USMC population.

(MOS 08xx) have risks for accidental deaths that are 37-percent to 69-percent higher than the 25 percent of the Marine Corps in the other omitted MOS categories. Notice that although these results are generally consistent with the picture presented in figure 4 that was based on fatality rates, there are differences that result from the correction for multiple risk factors.

Marines living on base have higher risk

To examine whether risk varies between those who live on base and those who live off, we used data on whether the Marine receives the Basic Allowance for Housing (BAH). Marines who are not eligible for BAH live in government-provided quarters and have a 21-percent higher risk than those living off base. This variable may reflect some underlying characteristics of barracks life (somewhat confining) that may contribute to a tendency to travel more. It may also reflect other individual characteristics associated with increased risks, such as being younger.

Marines assigned to rural commands have higher risk

Urban and rural areas may present very different risk factors, particularly for automobile accidents. To examine these effects, we include an urban variable. Marines who were assigned to San Diego (Miramar, Camp Pendleton, and MCRD San Diego), the Washington DC area (including Quantico), or Hawaii were considered to be in urban areas. The coefficient on the urban variable is .81 and statistically significant. This indicates that Marines in less congested area are about 23-percent more likely to suffer accidental deaths than Marines in these urban areas.

Early career has low risk followed by high risk

Recruit training is a restrictive environment that limits the opportunities for individuals to engage in risky behavior. We would have liked to estimate the risks associated with boot camp, but could not. The most troublesome issue is that there is incomplete demographic data on individuals who wash out (or die) during boot camp. This left us an incomplete picture of comparative risks for the first three months of service. However, we were able to investigate early career risk in subsequent periods.

We measure service relative to a Marine's active duty base date (roughly, the date of entry into active duty). One variable is used for the period 91 to 182 days after date of entry into service. A second is used for the next 6 months of active duty. The first of these corresponds to a period of generally close supervision and limited freedom. We find the risk of death is significantly lower during this period than at other times—about 40 percent of the risk faced in the time periods outside of the first 91 to 365 days in the Marine Corps. Conversely, the second half of Marine's first year is associated with much higher risk—over twice the risk associated with times outside these first 91 to 365 days.

E1s to E3s have higher risk

We include variables for three pay grade groups—E1 to E3, O1 to O3, and a group that combines senior officers and senior enlisted (O4 to O9, E7 to E9, and warrant officers). Thus, the coefficients represent the difference in risk for these groups relative to enlisted personnel in grades E4 to E6. The coefficients indicate that junior enlisted personnel have a 17 percent higher risk than the intermediate enlisted (but this is not statistically significant). The coefficient on junior officers indicates a much lower risk of death.

Reporting to a new command is associated with lower risk

We included variables to capture the risk associated with time frames relative to reporting to a new command, which we identified as entering a new present monitored command code (PMCC). We use one variable to cover the first 91 days at a new PMCC and another to cover the second 91 days. The first variable has a statistically significant coefficient of .33; the second variable is insignificant. This means that during the first 3 months at a new command, the risk of an accidental death is 33 percent of the risk for people who have been at a command for more than 6 months. This variable is correlated with the early career, because initial training is associated with frequent transfers among commands. However, the fact that both this and the second 3 months in the Marine Corps show significantly reduced risk suggests that both are separately meaningful.

Temporary duty is associated with lower risk

An individual who is assigned temporary additional duty (TAD) is at about 70 percent of the risk of a fatal accident as is an individual at his permanent duty station.

Promotions lower risk, but demotions are not significant

Some events might trigger specific behaviors. We examined certain career events to determine what was associated with higher risk. We include a variable for the 3 months after an individual was promoted. The coefficient indicates that the risk is only 59 percent of that for people with the same demographics who did not get promoted. Because people enlist at different ages and our analysis compares risks for people of a given age, this does not represent the difference in risk for a cohort that joins at a single time. That is, the coefficient is not reporting the risk for those who are promoted with those who were eligible and not promoted; rather, it is reporting the risk of those who were promoted compared to other Marines of the same age, whether they were eligible for promotion or not. (We do not have sufficient data to evaluate the risk associated with not being promoted when eligible.) Our promotion variable depends on the date a promotion was effective, not the date on which the individual learns he will be promoted.

We also evaluated the effects of demotions. We examined the risk of individuals in the 3 months immediately after they are demoted, and in the period between 3 and 6 months after demotion. The coefficients indicate no extra risk in the first 3 months, but over the next 3 months the risk of accidental death is more than doubled.

Return from deployment increases risk

Individuals who have recently returned from deployment appear to be at somewhat higher risk of accidents than those who have not recently returned. The hazard rate for people who are within 3 months of returning is a statistically insignificant 1.33. However, in the next 3 months, the hazard rate is elevated to 1.54, and is significant.

We should point out again that the deployment data are not of the highest quality. Often, the deployment return date is a projected date, which may be updated in later files. For those individuals who die soon after deployment, the correct date might not be available, because they may not be listed in any subsequent HMF files. Thus, it is possible that there is measurement error in this variable.

Enlistment waivers are often associated with higher risk

Individuals with a history of drug use, traffic violations, or serious offenses require waivers to enter the Marine Corps. Each of these factors is associated with a 66- to 83-percent increase in risk.[7] However, Marines who entered with waivers for misdemeanors have 33-percent less risk.

Individuals who did not complete high school have lower risk

We included variables for people whose education went beyond high school and those who did not have a high school diploma. The coefficient for people who did not complete school was significant and indicated a reduced risk of 33 percent, relative to those who completed high school but had no further education. The coefficient on education beyond high school was not significant. Differences in AFQT test scores are not associated with higher or lower risks.

Seasonal effects

We included variables to represent the quarters of a calendar year. The results indicate that fatal accidents are least likely to occur in the January–March quarter. They are most likely to occur in the July–September quarter, when the risk is 52-percent more than in January–March. The March–June and October–December quarters are also higher risk quarters, with a 32- to 34-percent increase over January–March. An explanation may be that less travel takes place in the winter quarter.[8] The seasonal effects are similar to those observed nationwide.

---

7. We caution against concluding that these results suggest changing recruitment policy. Large numbers of recruits do enter under waivers and earlier CNA work has shown that they are often good Marines [3, 4].

8. Since we use individual data, other seasonal factors related to the recruitment cycle (e.g., age mix, number of Marines) are already implicitly accounted for.

Combining risk factors

The above hazard rates provide information about how risk varies with a single characteristic. However, there might be interest in determining how risk changes with multiple characteristics. To estimate the relative risk between two people who differ by multiple characteristics, we can use the coefficients in the second column of table 2. Specifically, the value of each variable that differs between the individuals is multiplied by the corresponding coefficient, the result is summed, and the exponential function is applied to that result.

As an example, suppose we wanted to estimate the risk for a black male aviation mechanic as compared with a white male in one of the MOSs not listed. (All other characteristics are assumed to be the same for the two individuals.) We can calculate the relative risk as

$$\begin{aligned}
&= \exp(\beta_{black} black + \beta_{MOS60} MOS60) \\
&= \exp(0.300 + .440) \\
&= 2.1
\end{aligned}$$

The coefficients are drawn from table 3 (for convenience, pertinent lines from table 2 are repeated here in table 3). The variables black and MOS60 are both 1, because they are indicator variables (i.e., equal to 1 if the individual has the characteristic and equal to 0 otherwise). Notice that we did not have to include a variable for male because the comparison is between two males. We did not include other variables, because they too are assumed to be the same for both individuals. Once the relevant values are specified, we can calculate the hazard rate. In this case, the combined risk is 2.1 times that of the reference individual.

Table 3. Selected estimation results from table 2

	(1) Hazard ratio	(2) Coefficient	(3) p-value
Black	1.35	0.300	0.023
Male	1.53	0.417	0.100
Aircraft maintenance (MOS 60xx)	1.55	0.440	0.049

Estimating risks for vehicle-related deaths

Because fatalities related to motor vehicles account for about 70 percent of the accidental deaths, and the characteristics associated with motor vehicle accidents might differ from those of other accidents, we performed a separate analysis using motor vehicle deaths as the event of interest.

The estimation results are listed in table 4. There are a number of differences in how the various characteristics are associated with motor-vehicle deaths.

Table 4. Estimation results for the risk of motor-vehicle related death[a]

	Hazard ratio[b]	Coefficient	p-value
Male	1.26	0.23	0.380
Urban (DC, San Diego/Pendleton, Hawaii)	0.83	-0.19	0.104
Personnel and admin (MOS 01xx)	0.85	-0.16	0.529
Infantry (MOS 03xx)	1.02	0.02	0.904
Logistics (MOS 04xx)	0.77	-0.26	0.545
Command and control systems (MOS 06xx)	1.08	0.08	0.812
Field artillery (MOS 08xx)	1.14	0.13	0.697
Engineers (MOS 13xx)	1.56	0.45	0.050**
Ordnance (MOS 21xx)	1.20	0.18	0.607
Data/communications maintenance (MOS 28xx)	1.04	0.04	0.916
Supply admin and operations (MOS 30xx)	0.97	-0.03	0.901
Motor transport (MOS 35xx)	1.36	0.31	0.129
Military police and corrections (MOS 58xx)	1.03	0.03	0.932
Aircraft maintenance (MOS 60xx)	1.41	0.34	0.177
Helicopter maintenance (MOS 61xx)	1.21	0.19	0.520
Avionics (MOS 63xx)	1.21	0.19	0.580
Pilots (MOS 75xx)	1.16	0.14	0.844
Special identifier MOSs (MOS 99xx)	1.01	0.01	0.983
Hispanic	1.30	0.26	0.091*
Black	1.38	0.32	0.033**
Other non-white races	1.00	0.00	0.997
Physical fitness test – fail (class 4)	0.87	-0.14	0.752
Physical fitness test – low pass (class 3)	0.63	-0.47	0.081*
Physical fitness test – med pass (class 2)	0.95	-0.05	0.678
Junior enlisted (E1-E3)	1.20	0.19	0.222
Senior officers and enlisted (O4-O9 & E7-E9))	0.49	-0.71	0.244
Junior officers (O1-O3)	0.54	-0.62	0.214
91–182 days since entry into USMC	0.44	-0.81	0.036**
182–365 days since entry into USMC	2.32	0.84	0.000***

Table 4. Estimation results for the risk of motor-vehicle related death[a] (cont'd)

	Hazard ratio[b]	Coefficient	p-value
0–91 days since reporting to command	0.32	-1.13	0.000***
91–182 days since reporting to command	0.76	-0.27	0.177
0–91 days since promotion	0.54	-0.62	0.001***
0–91 days since demotion	1.27	0.24	0.542
91–182 days since demotion	2.45	0.90	0.007***
Married or with dependents	0.68	-0.38	0.003***
On temporary duty (TAD)	0.70	-0.35	0.081*
0–91 days since returning from deployment	1.84	0.61	0.024**
91–182 days since returning from deployment	2.10	0.74	0.006***
Education beyond high school	0.66	-0.41	0.271
Less than a high school education	0.61	-0.50	0.058*
AFQT score	1.00	0.00	0.563
AFQT score not available	1.34	0.29	0.464
Waiver for serious offense	1.61	0.48	0.016**
Waiver for misdemeanor	0.69	-0.37	0.184
Waiver for drugs	2.08	0.73	0.005***
Serious offense waiver not available	0.98	-0.02	0.914
Waiver for traffic violations	1.58	0.46	0.261
Living on base	1.21	0.19	0.143
April–June	1.52	0.42	0.018**
July–September	1.62	0.48	0.003***
October–December	1.68	0.52	0.001
Quarterly time trend	1.00	0.00	0.868
Wald Chi-squared test statistic	190.01		0.000***

a. * significant at 10%; ** significant at 5%; ***significant at 1%

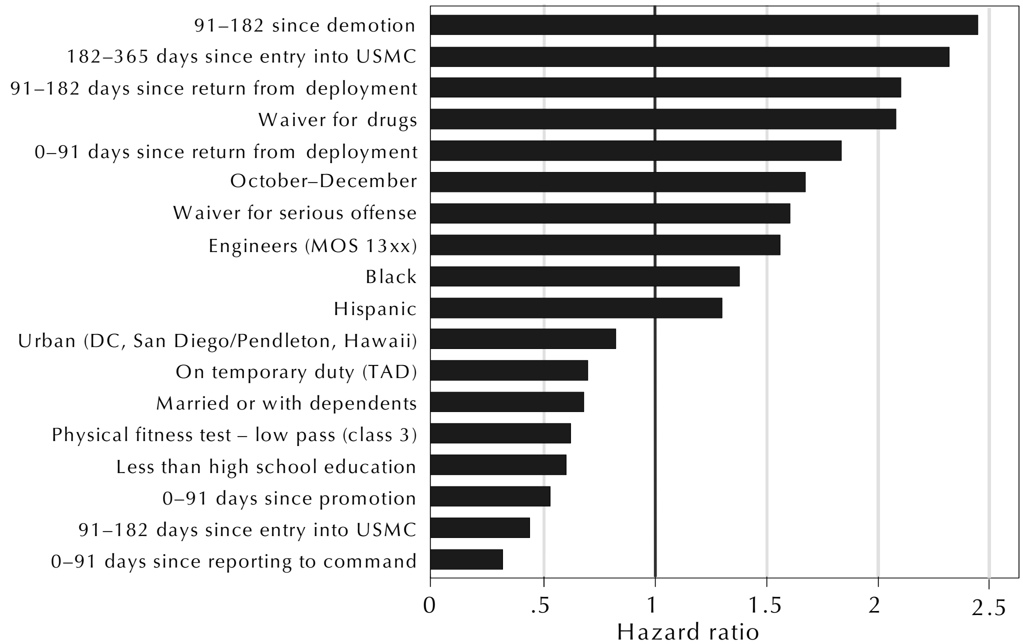
b. For categorical variable, hazard ratios are interpreted relative to an excluded category. For example, the Black and Hispanic values are relative to whites and MOS values are relative to individuals in MOS groups not listed.

In the following sections, we highlight results that are likely to be of most interest. Perhaps because we are now using a smaller number of deaths, some variables that were previously statistically significant are no longer significant. Figure 9 provides a summary of the significant results.

No significant differences between male and female

The coefficient on males was an insignificant 1.26, suggesting that males may be at higher risk than females, but not at a statistically significant level.

Figure 9. Vehicle fatalities, risk associated with various characteristics



Most variables had similar effects

Not surprisingly, many of the variables had coefficients and significance levels that were very similar to the hazard rates estimated for accidental deaths. Included in this category are

- Race
- Married or with dependents
- Physical fitness test (PFT)
- Junior enlisted
- Early career
- Reporting to new command
- Promotion and demotion
- Temporary duty (TAD) status
- Education
- Living on base (the coefficient is only significant at 14 percent)

- Urban (the coefficient was similar, but it is now borderline significant at the 10-percent level).

Engineer MOS is at higher risk

The only statistically significant coefficient among the 16 largest MOSs is associated with the engineering/construction MOS, which has a 56-percent increase in risk. A number of other variables were similar to the model of all deaths, but were no longer significant. In particular, the Motor Transportation MOS (MOS 35xx) hazard rate dropped from 1.55 to 1.36, Aviation Maintenance MOS (MOS 60xx) hazard rate dropped from 1.55 to 1.41.

Other MOSs showed high hazard ratios in the model of all deaths but were not significant in this model of motor vehicle deaths. In particular, the infantry MOS (MOS 03xx) and field artillery MOS (MOS 08xx) dropped from a relatively high hazard ratios to near 1, with insignificant p-values. This is not entirely surprising. We saw in figure 4 that vehicle-related deaths represented a smaller share of fatalities for these two groups than for other Marines.

Return from deployment is significant

Whereas the effect of returning from deployment was not strong in the model of all accidents, it is very strong in the motor vehicles model. The first 3 months after a deployment have an 84-percent increased risk, and the next three months have a 110-percent increase.

Traffic waiver is large but insignificant

The risk of accidental motor-vehicle death associated with entering the Marine Corps with a waiver for traffic violations is 58-percent higher than for those who did not enter with such a waiver. However, the significance of this variable is fairly low, with a p-value of 0.261. This is a surprising result, especially because the variable is larger and is significant in the model based on all accidental deaths.

Waivers for serious offenses and for drug offenses were associated with significantly higher risk of fatal motor-vehicle accidents, as they were for all fatal accidents.

Time of year

The time of year risks increase relative to the winter quarter. The risk of a fatal motor-vehicle death in the October–December quarter is statistically significant and is 68-percent higher than in the January–March quarter. This is much higher than the earlier model indicated. The March–June and October–December quarters are also high risk quarters.

Predictions, using a parametric model

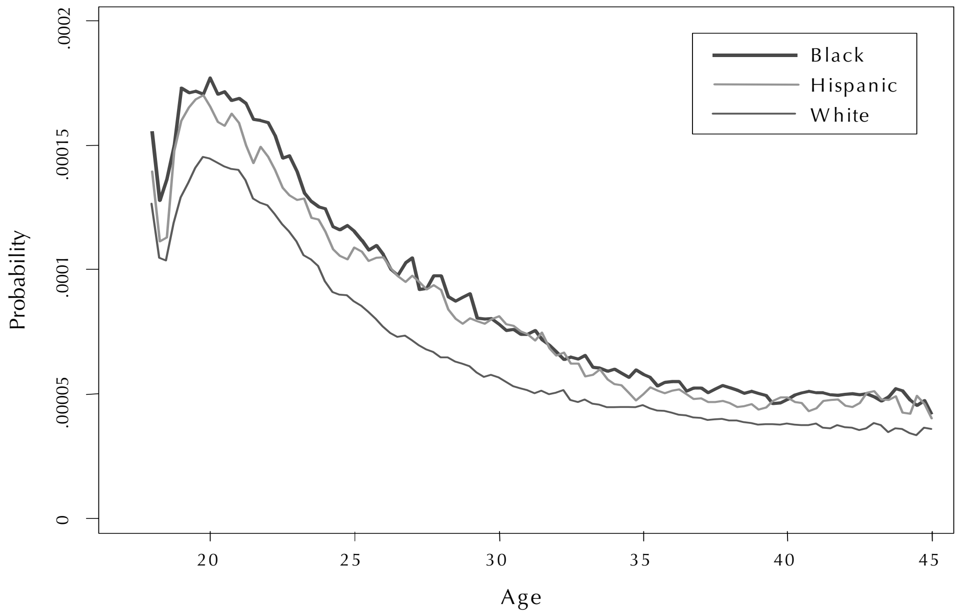
Earlier, we commented briefly on a linear regression approach to prediction based on aggregate fatality data by quarter. Our use of individual data gives us the ability to reflect individual risk and allows for richer predictions than are possible with typical regression models. In particular, we can deal with situations where there may be a substantial change in the underlying mix of risk factors (e.g., large numbers of Marines returning from deployment).

In the analysis so far, we did not need to specify how baseline risk varied with age. However, we can more easily generate predictions if we use a *parametric hazard model,* one that specifies a particular functional form for the baseline hazard function $h_0(t)$. Specifically, we will now assume that $h_0(t)$ follows a Weibull function. The Weibull was selected for its flexibility and fit to the apparent shape of baseline hazard function revealed by our previous model. The individual coefficients we estimate with this new model are very similar to those already presented in table 2 (see table 5 in appendix B).

Using the parametric model to estimate probability of death

Figure 10 presents the estimated probability of accidental death by race and age. These estimates incorporate the actual mix of demographics by race and age. That is, for a given race and age group, it shows the mean probability of death, reflecting the combination of individual characteristics (such as concentrations in MOSs and career status) within the current Marine Corp population. The figure suggests that each group has a similar pattern of generally decreasing risk with age, but it also shows a higher overall risk for blacks and Hispanics.

Figure 10. Estimated probability of accidental death, by age and race



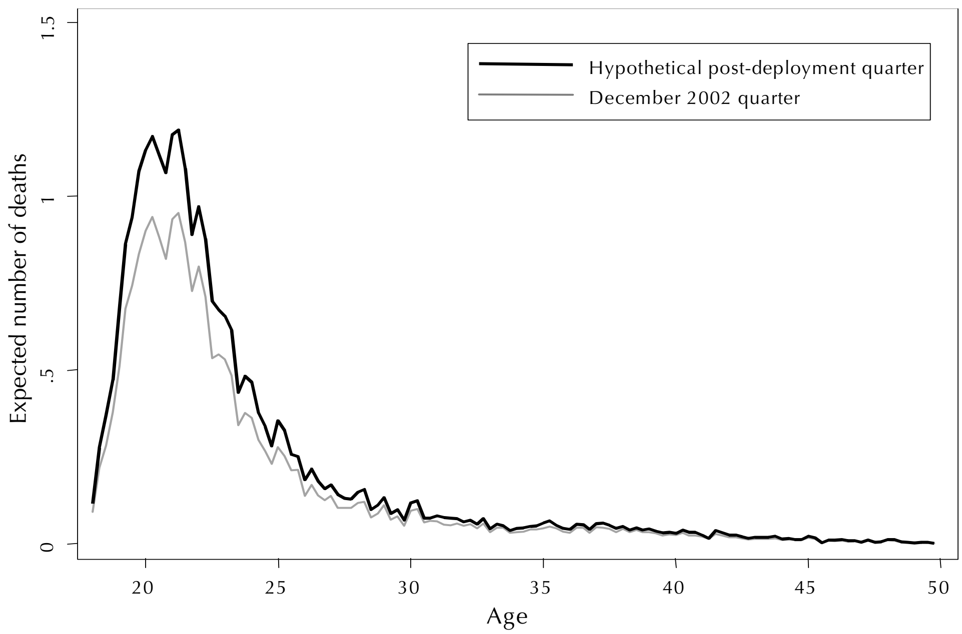
Using the parametric model to predict fatalities

Again using the parametric model, a predicted number of overall deaths can be generated. This prediction will reflect the characteristics of each individual in the Marine Corps. Figure 11 presents the expected number of deaths by age for the quarter ending in December 2002 (the lower line in the figure). We compute this number by summing the individual probabilities of death (probability of a fatal accident on single day) for each Marine and multiplying by 91 to get the quarterly total. The total number of deaths can be found as the sum across the age groups. The total number of deaths predicted is 21 (compare this to the results of the regression fit of quarterly data shown in figure 3).

Predictions can account for changes in underlying risk

Now suppose half the Marine Corps were to return from a deployment at one time. That would present a challenge for standard predictions methods based on aggregate fatality data, but the situation is easily handled in a model that accounts for individual probability of death. Based on our analysis, we would expect a significant increase in deaths in the period spanning 91- to 182-days after this hypothetical return from deployment. Figure 11 presents the predicted number of deaths by age during that 3-month period (upper line). For comparison, a more typical December quarter is also shown. The total expected number of deaths is now 27, up from 21 otherwise expected.

Figure 11. Predicted number of deaths per quarter, by age



Appendix A: Specification tests

We checked the specification of the estimated models in several ways. First, as reported in tables 2 and 4, the Wald statistic is large, indicating that the models as a whole are statistically significant in explaining deaths. Second, we tested the linear specification of explanatory variable (linear within the exponential hazard term). We did so by estimating a hazard model for deaths based on the predicted linear combination $x\hat{\beta}$ and the square of this linear combination. The coefficient on the squared term was insignificant, indicating that the linear specification is suitable. Finally, we tested the proportional hazards assumption. For the model of all accidental deaths, this assumption cannot be rejected based on a test of the slope of the residuals, nor can it be rejected for either the combined set of variables or any individual variable. For the model of vehicle-related deaths, the assumption of proportional hazards cannot be rejected for the combined set of variables, but 3 of the 52 individual variables were rejected (MOS06, MOS08, and 191–182 days since demotion).

This page left intentionally blank.

Appendix B: Parametric model results

Table 5. Parametric model results for the risk of accidental death[a]

	(1) Hazard ratio[b]	(2) p-value
Male	1.53	0.092*
Urban (DC, San Diego/Pendleton, Hawaii)	0.82	0.047**
Personnel and admin (MOS 01xx)	0.88	0.594
Infantry (MOS 03xx)	1.37	0.028**
Logistics (MOS 04xx)	0.86	0.675
Command and control systems (MOS 06xx)	1.05	0.882
Field artillery (MOS 08xx)	1.68	0.057*
Engineers (MOS 13xx)	1.58	0.029**
Ordnance (MOS 21xx)	1.36	0.321
Data/communications maintenance (MOS 28xx)	0.93	0.811
Supply admin and operations (MOS 30xx)	1.15	0.523
Motor transport (MOS 35xx)	1.56	0.013**
Military police and corrections (MOS 58xx)	1.14	0.675
Aircraft maintenance (MOS 60xx)	1.54	0.053*
Helicopter maintenance (MOS 61xx)	1.29	0.332
Avionics (MOS 63xx)	1.01	0.968
Pilots (MOS 75xx)	1.79	0.212
Special identifier MOSs (MOS 99xx)	1.35	0.234
Hispanic	1.25	0.100*
Black	1.34	0.028**
Other non-white races	0.87	0.567
Physical fitness test - fail (4)	0.82	0.640
Physical fitness test - low pass (3)	0.68	0.091*
Physical fitness test - med pass (2)	0.95	0.627
Junior enlisted (E1 to E3)	1.21	0.141
Senior enlisted (E7 to E9) & senior officers (O4 to O9)	0.71	0.229
Junior officers (O1 to O3)	0.44	0.070*
91–182 days since entry into USMC	0.40	0.007***
182–365 days since entry into USMC	2.00	0.000***
0–91 days since reporting to command	0.32	0.000***
91–182 days since reporting to command	0.79	0.152
0–91 days since promotion	0.59	0.001***
0–91 days since demotion	0.94	0.873
91–182 days since demotion	2.03	0.023**
Married or with dependents	0.67	0.000***

Table 4. Parametric model results for the risk of accidental death[a] (cont'd)

	(1) Hazard ratio[b]	(2) p-value
On temporary duty (TAD)	0.71	0.045**
0–91 days since returning from deployment	1.33	0.270
91–182 days since returning from deployment	1.57	0.084*
Education beyond high school	0.88	0.629
Less than a high school education	0.67	0.069*
AFQT score	1.00	0.310
AFQT score not available	1.33	0.401
Waiver for serious offense	1.65	0.004***
Waiver for misdemeanor	0.66	0.082*
Waiver for drugs	1.67	0.042**
Serious offense waiver not available	0.90	0.456
Waiver for traffic violations	1.82	0.073*
Living on base	1.21	0.105
April–June	1.31	0.076*
July–September	1.51	0.003***
October–December	1.35	0.024**
Quarterly time trend	1.01	0.059*
Wald Chi-squared test statistic	213.89	0.000***

a. * significant at 10%; ** significant at 5%; ***significant at 1%.

b. For categorical variable, hazard ratios are interpreted relative to the excluded category. For example, the Black and Hispanic values are relative to whites and MOS values are relative to individuals in MOS groups not listed.

References

[1] Nicholas M. Kiefer. "Economic Duration Data and Hazard Functions." *Journal of Economic Literature*, 26, Jun 1988, pp 646-679

[2] Mario A. Cleves, William W. Gould, and Roberto G. Gutierrez. *An Introduction to Survival Analysis Using Stata*. College Station, Texas: Stata Press, 2002

[3] Aline Quester. *Meritorious Promotions and Enlistment Waiver Status*, Jul 2001 (CNA Memorandum D0004311.A1)

[4] Aline Quester. *Enlistment Waivers, Educational Background, and Attrition From the Marine Corps: Male Accessions*, Jan 1999 (CNA Memorandum 98-1619)

This page left intentionally blank.

List of figures

This page left intentionally blank.

List of tables

This page left intentionally blank.

CRM D0008914.A2/Final